Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

Career Pattern Analysis of SMKN 1 Stabat Graduates Using K-Means Clustering Algorithm on Tracer Study Dataset

Ibrahim[1], Muhammad Iqbal[2]
[1,2]Pascasarjana, Magister Teknologi Informasi, Universitas Pembangunan Pancabudi, Medan, Indonesia

Article Info
Keywords:
Tracer Study, K-Means Clustering, Rapidminer, Career Pattern, Graduates, Data Analysis

ABSTRACT
Tracer study is a method commonly used to determine the condition of graduates of an educational institution, including the career patterns they pursue. This study aims to analyze the career patterns of SMKN 1 Stabat graduates by utilizing the K-Means clustering algorithm. The dataset was obtained from the results of a tracer study of 287 alumni of SMKN 1 Stabat. The dataset used came from a tracer study conducted on graduates in the last five years. By grouping data using K-Means, it is hoped that specific patterns can be found that can help schools improve the quality of learning and student work readiness.[4] The results of the analysis show several dominant career pattern groups, such as the industrial sector, entrepreneurship, and further education.

Corresponding Author:
Ibrahim,
Pascasarjana, Magister Teknologi Informasi,
Universitas Pembangunan Pancabudi, Medan, Indonesia
ibrahim31@guru.smk.belajar.id

INTRODUCTION

Vocational High Schools (SMK) have an important role in preparing skilled workers who are ready to enter the workforce. SMKN 1 Stabat, as one of the vocational education institutions in North Sumatra, continues to strive to improve the quality of its graduates to meet the needs of industry and the job market. One way to evaluate the success of an education program is through a tracer study, which provides an overview of the career patterns and professional development of alumni. Tracer studies are an important method for evaluating educational outcomes and their relevance to the world of work.[15] Vocational education tracer studies, hereinafter referred to as tracer studies, are surveys to determine work activities (working, entrepreneurship and continuing education), alignment, and satisfaction with the world of work for vocational education graduates after one year of graduating from vocational education units. For Vocational High Schools (SMK), tracer studies can provide valuable feedback on the quality of graduates and the suitability of their competencies to industry needs.[3][19]

In the era of big data, analysis of tracer study datasets can be done more effectively using data mining techniques, especially clustering algorithms.[5] This method allows alumni data to be grouped based on their career characteristics, so that emerging patterns can be identified. The results of this analysis are very valuable for schools to evaluate and improve the curriculum, as well as design more targeted programs to improve graduates' work readiness.[2]In the digital era, it is important for educational institutions to understand

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

the career patterns of their graduates. This information is not only an indicator of the success of the institution, but also the basis for developing a curriculum that is more relevant to the needs of the job market. SMKN 1 Stabat as one of the vocational high schools that focuses on preparing students for the world of work, requires a data-based approach to evaluate its educational outcomes. The use of the K-Means algorithm in this study is very effective in grouping complex tracer study data into more easily understood information. Through the clustering process, data such as employment status, industrial sector, and further education can be grouped into four main clusters. Each cluster reflects the characteristics of different career patterns and provides a clear picture of the distribution of graduates. In addition, the career patterns that have been successfully identified can be used as evaluation material for SMKN 1 Stabat in formulating curriculum development strategies. For example, the dominance of graduates in the industrial sector indicates the need to strengthen practical skills and cooperation with companies. Meanwhile, for graduates who choose to become entrepreneurs, a more intensive and practice-based entrepreneurship program can be implemented.

The evaluation process of clustering results using silhouette scores ensures the quality of data grouping.[9] Silhouette score values approaching 1 indicate that the data in the cluster have high similarity, so that the clustering results can be trusted to be used in decision making. This proves the reliability of the K-Means algorithm in analyzing tracer study data.[4] [6] Furthermore, this study also opens up opportunities for future research development. The use of other machine learning algorithms such as DBSCAN or Hierarchical Clustering can be explored to provide a comparison of results.[16] In addition, broader data collection with additional attributes can provide a more detailed picture of graduate career patterns. Overall, the results of this study indicate that career pattern analysis with K-Means provides significant benefits for the development of school policies.[20] By understanding graduate career patterns, SMKN 1 Stabat can develop learning programs that are more effective and responsive to the demands of the world of work and student needs.

RESEARCH METHODOLOGY

The research stage flow is a series of systematic steps that must be passed by a researcher in carrying out this research methodology involves several systematic stages to ensure valid and accurate analysis results. The following is an overview of the research stage flow:

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

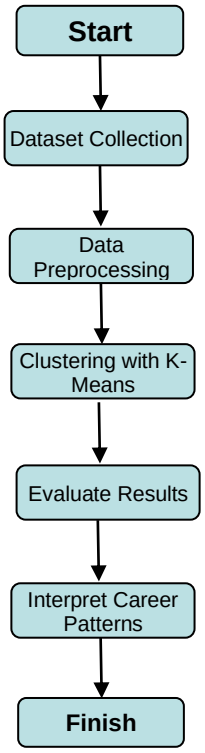


Figure 1.Research Stages

This image is a research flow diagram consisting of several main stages:

1 Start Research
The research begins by determining the main focus, namely the analysis of career patterns of SMKN 1 Stabat graduates through tracer study data. This stage involves compiling the background of the problem, formulating objectives, and compiling a research framework.

2 Identification of problems
This step aims to identify the main problems to be studied, such as the gap between the field of study and the world of work, and the need for grouping career patterns of graduates. The formulation of the problem becomes a guide in designing the research methodology.

3 Data collection
Tracer study data was collected through questionnaires distributed to alumni of SMKN 1 Stabat. The information collected includes the type of work, industry, work location, and the relevance of the work to the field of study. In addition, the data was supplemented with structured interviews to obtain a deeper context.

4 Dataset Preparation
The collected data is validated and compiled into a dataset ready for analysis. Validation is done to ensure the completeness and consistency of the data, so that the analysis results can be relied upon.

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

5 Analysis with K-Means
The K-Means algorithm is applied to the dataset to cluster graduate career patterns based on certain characteristics. This process includes determining the optimal number of clusters, grouping the data, and visualizing the analysis results.

6 Interpretation of Results
The clustering results are analyzed to identify key patterns emerging in the data. Key findings are interpreted by relating them to relevant theory and previous research.
Factors influencing clustering are also described.

7 Preparation of Reports
The research findings are compiled in the form of a report that includes background, methods, results, and discussion. This report also includes strategic recommendations for schools in improving the relevance of the curriculum and student career development.

8 Finished
The research is completed after the report is prepared and submitted to the relevant parties. The results of the research are expected to provide practical benefits for the development of vocational education at SMKN 1 Stabat.

This flowchart shows the research steps systematically from problem identification to the preparation of the final report.

Dataset

The dataset used in this study comes from a tracer study conducted on SMKN 1 Stabat graduates in the last five years. This data includes various important attributes such as expertise programs, employment status, employment sectors, income, work location, and continuation of studies. This dataset was chosen because of its relevance in describing graduate career patterns comprehensively.

Data Preprocessing

The preprocessing stage aims to prepare the data so that it is ready to be used in the analysis. The steps taken include:

a. Data Cleansing: Delete or replace missing values so that they do not affect the clustering results.
b. Data Transformation: Converts categorical data, such as employment sector, to numeric format using encoding.
c. Data Normalization: Standardize the range of attribute values to have balanced weights, so that no attribute dominates the clustering process.

K-Means Algorithm

Is one of the unsupervised learning methods used to group data into several groups (clusters) based on similar characteristics. This algorithm works by dividing data into k groups, where k is the number of groups determined in advance.[14][7]

The K-Means algorithm is used to cluster the dataset. This process involves the following steps:

1. Initialize the Number of Clusters (K): Determine the initial number of clusters based on

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

data exploration.

2. Determination of Initial Centroid: The initial centroid is randomly selected from the data.
3. Grouping Iteration: Each data is allocated to a cluster based on the closest distance to the centroid. The centroid is then updated based on the average of the data in the cluster.
4. Convergence: The process is repeated until the centroids no longer change significantly.

Evaluation of Clustering Results

Evaluation of clustering results is done using the silhouette score method, which measures how well data in a particular cluster is grouped compared to other clusters. Silhouette score values range from -1 to 1, where values close to 1 indicate good clustering. Sample data after pre-processing is as follows:

Table 1.sample data taken from Google Form BKK SMKN 1 Stabat

NO	Student Name	Graduation year	Expertise Competence	Gender	Status
1	Two Marlina	2019	FASHION	Woman	Studying
2	MAJUS TIYA WATI	2019	FASHION	Woman	Work
3	NURSING AGUSTIWI	2021	FASHION	Woman	Work
4	Erni Sidauruk	2019	FASHION	Woman	Work
5	NURSING AGUSTIWI	2021	FASHION	Woman	Work
6	NATALIA BR. SINAGA	2019	FASHION	Woman	Businessman
7	Sundari Sartika	2021	FASHION	Woman	Businessman
8	Yulia Sari	2021	FASHION	Woman	Work
9	NIKEN ARISNA	2020	FASHION	Woman	Studying
10	Ricky Syahputra	2021	Computer Network Engineering (TKJ)	Man	Work
11	Ike Padilla	2021	Computer Network Engineering (TKJ)	Woman	Studying
12	Adilla Azzahra	2020	FASHION	Woman	Businessman
13	SAQINA ARDANA	2020	FASHION	Woman	Businessman

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

NO	Student Name	Graduation year	Expertise Competence	Gender	Status
14	HILDA TRIANANTA	2021	FASHION	Woman	Businessman
15	Win Gomgom Parsaulian Sirait	2019	Computer Network Engineering (TKJ)	Man	Studying
16	THE WORD OF THE LORD	2021	FASHION	Woman	Work
17	Muhammad Ali Arsyad	2019	Machining Engineering (TP)	Man	Work
18	SRI MELATI	2021	ACCOUNTANCY	Woman	Not yet working
19	Nuril Mahzula Nasution	2021	ACCOUNTANCY	Woman	Studying
20	Fitri Awaliyah	2021	ACCOUNTANCY	Woman	Studying
21	Monica Aprilia	2021	ACCOUNTANCY	Woman	Work
22	INDONESIA	2020	ACCOUNTANCY	Woman	Businessman
23	MY DAY	2021	FASHION	Woman	Businessman
24	Muhammad Zacwan	2021	ACCOUNTANCY	Man	Work
25	RISMA	2020	ACCOUNTANCY	Woman	Businessman
26	Fitri andini	2021	ACCOUNTANCY	Woman	Work
27	The Star of Mala Sari	2021	ACCOUNTANCY	Woman	Work
28	NURIL MAHFUZA NASUTION	2021	ACCOUNTANCY	Woman	Studying
29	FATHIA AZAHRA	2020	FASHION	Woman	Work
30	CINDY PRAISKA	2021	ACCOUNTANCY	Woman	Studying

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

NO	Student Name	Graduation year	Expertise Competence	Gender	Status
31	DILLA FITALOKA	2020	ACCOUNTANCY	Woman	Work
32	Muhammad Ayub	2021	Machining Engineering (TP)	Man	Work
33	Prabowo	2021	Machining Engineering (TP)	Man	Work
34	Nada Ridona	2019	FASHION	Woman	Studying
35	Muhammad Nabil	2021	Machining Engineering (TP)	Man	Work
36	DEBBY FAUZIAH	2019	FASHION	Woman	Work
37	SABRINA ALMAIRA DAULAY	2021	MARKETING	Woman	Studying
38	Rizqina Masrura	2020	MARKETING	Woman	Work
39	Winda Windari	2021	MARKETING	Woman	working
40	EVI SYAHFITRI	2019	MARKETING	Woman	Studying
41	Amanda Khairunnisa Putri	2021	ACCOUNTANCY	Woman	Studying
42	SULISTIO	2021	Light Vehicle Engineering	Man	Work
43	Ade Rian Syahputra	2019	Light Vehicle Engineering	Man	Work
44	RIYAN SEYHAN NUR	2020	Light Vehicle Engineering	Man	Studying
45	AYUDRI PRINCESS PRAMITA	2020	MARKETING	Woman	Work

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

NO	Student Name	Graduation year	Expertise Competence	Gender	Status
46	THE GREEK HERO	2020	Computer Network Engineering (TKJ)	Man	Work

NO	Student Name	Graduation year	Expertise Competence	Gender	Status
47	M.ALPIQRIANSYAH	2018	Computer Network Engineering (TKJ)	Man	Work
48	Khairun Nazri	2020	Computer Network Engineering (TKJ)	Man	Work
49	Muhammad Hafiz	2021	Computer Network Engineering (TKJ)	Man	Studying
50	The story of Agum Dwi Syahputra	2020	Computer Network Engineering (TKJ)	Man	Work
51	Dwi Putri Fariza	2019	Computer Network Engineering (TKJ)	Woman	Studying
52	ELIA WATI	2018	Computer Network Engineering (TKJ)	Woman	Not yet working
53	Beautiful Sari	2018	Computer Network Engineering (TKJ)	Woman	Work
54	Khelfin light pratama	2021	Motorcycle Engineering (TBSM)	Man	Work
55	Nazwa Arifa	202	Computer Network Engineering (TKJ)	Woman	Businessman

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

NO	Student Name	Graduation year	Expertise Competence	Gender	Status
56	ALFIN FAUZI	2021	Light Vehicle Engineering	Man	Not yet working
57	Nazla Annisa Gunawan Sirait	2021	Computer Network Engineering (TKJ)	Woman	Studying
58	SONNI ANDRYAN HUTAPEA	2021	Computer Network Engineering (TKJ)	Man	Studying
59	Ridho Zumara	2020	Light Vehicle Engineering	Man	Studying
61	PRINCESS ANJELITA	2019	ACCOUNTANCY	Woman	Work
62	The Lord of the Worlds	2020	Computer Network Engineering (TKJ)	Woman	Work
63	DIAN SYAHPUTRA	2018	Computer Network Engineering (TKJ)	Man	Studying
64	Lily Rahmawati	2020	Computer Network Engineering (TKJ)	Woman	Studying
65	ALDO ARADHEA	2020	Motorcycle Engineering (TBSM)	Man	Studying
66	Dwi Syafitri	2019	OFFICE ADMINISTRATION	Woman	Work
67	Dila Ramadhani	2019	OFFICE ADMINISTRATION	Woman	Work
68	RIZCHA DIANTI	2020	OFFICE ADMINISTRATION	Woman	Work
69	RIANI	2019	OFFICE ADMINISTRATION	Woman	Studying
70	Kartika	201	OFFICE ADMINISTRATION	Woman	Work

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

NO	Student Name	Graduation year	Expertise Competence	Gender	Status
	Handayani	9			
71	Nurlela Sari	2018	OFFICE ADMINISTRATION	Woman	Studying
72	The Story of Marshanda Febriyanti	2018	OFFICE ADMINISTRATION	Woman	Work
73	DILA WANDA SARI	2019	OFFICE ADMINISTRATION	Woman	Businessman
74	Tarissah	2019	OFFICE ADMINISTRATION	Woman	Work
75	The sun is shining	2019	OFFICE ADMINISTRATION	Woman	Work
76	Mega Cynthia	2018	OFFICE ADMINISTRATION	Woman	Work
77	Leni Apriliya	2020	OFFICE ADMINISTRATION	Woman	Studying
78	Legitimate	2019	OFFICE ADMINISTRATION	Woman	Work
79	Zidan Fadli Syahputra	2021	Motorcycle Engineering (TBSM)	Man	Not yet working
80	SITI MAYSARAH	2021	Computer Network Engineering (TKJ)	Woman	Not yet working
			Computer		

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

NO Student Name	Graduation year	Expertise Competence	Gender	Status
81	RIDHA PRAMA MITA	2020	Network Engineering (TKJ)	Woman	Studying
....	.....	.....	.....	.....	.....
....	.....	.....	.....	.....	.....
287	ICHAH PRATIWI	2024	ACCOUNTANCY	Woman	Work

Figure 2.Data on the Number of Tracer Study Students at SMKN 1 Stabat

RESULTS AND DISCUSSION

Data Input

Below is an image that shows the data that will be Clustered in Rapidminer, namely the first stage of Operator Selection.

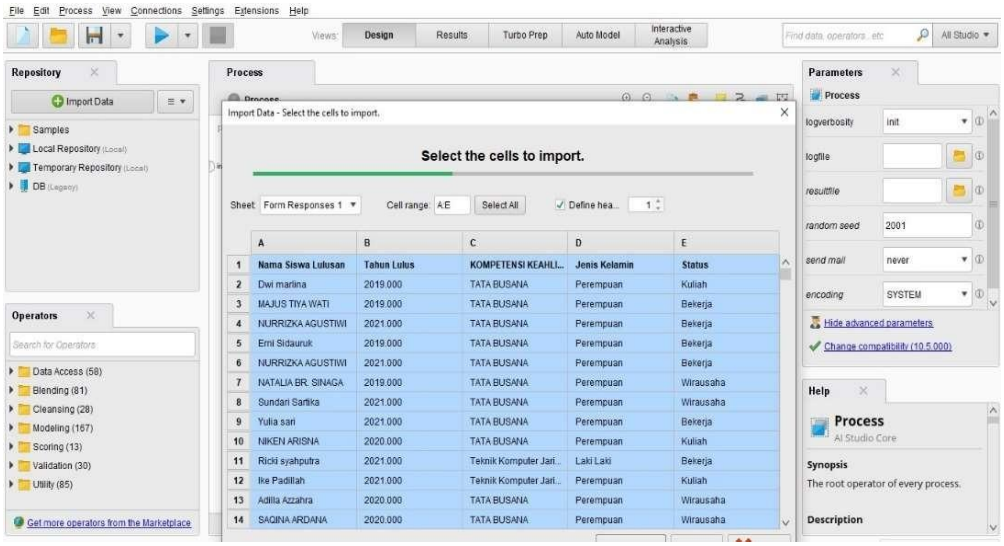


Figure 3. Selection

Data Preprocessing Stage

The next stage is data preprocessing, which includes several important steps, such as removing duplicate data, correcting errors in the data, and other relevant steps. The main goal of this stage is to produce clean and quality data attributes, so that the data is ready to be used for the next process, namely data transformation.

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

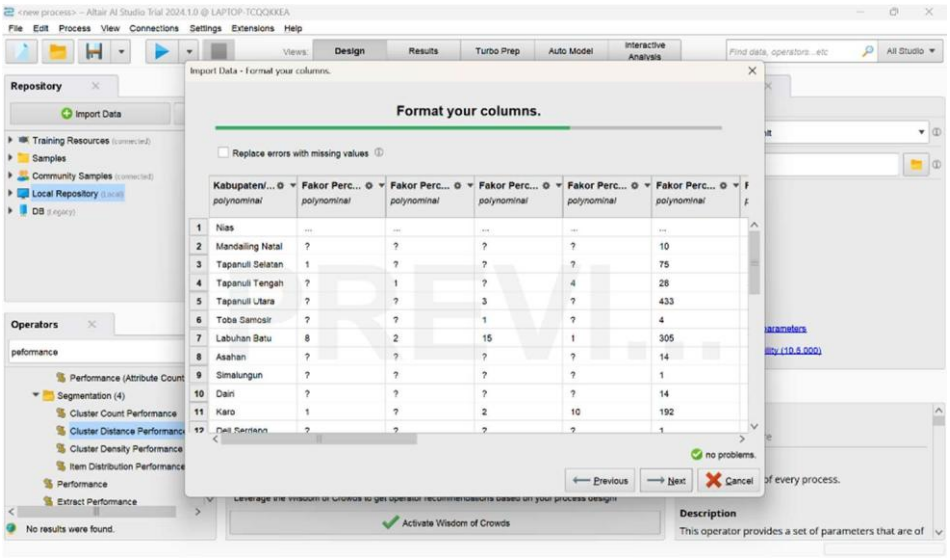
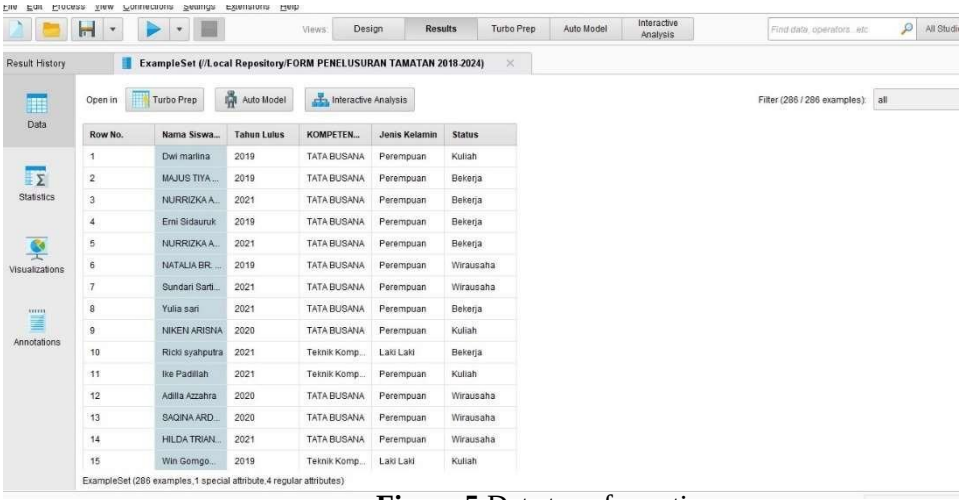


Figure 4. Preprocessing data

Data Transformation

The transformation stage is a process in which data that has gone through the previous stage is transformed into a format or form that suits the researcher's needs. At this stage, the data is prepared to be used effectively in the next steps in the research process.



Row No.	Nama Siswa...	Tahun Lulus	KOMPETEN...	Jenis Kelamin	Status
1	Dwi marlina	2019	TATA BUSANA	Perempuan	Kuliah
2	MAJUS TIYA...	2019	TATA BUSANA	Perempuan	Bekerja
3	NURRIZKA A...	2021	TATA BUSANA	Perempuan	Bekerja
4	Erni Sidauruk	2019	TATA BUSANA	Perempuan	Bekerja
5	NURRIZKA A...	2021	TATA BUSANA	Perempuan	Bekerja
6	NATALIA BR. ...	2019	TATA BUSANA	Perempuan	Wirausaha
7	Sundari Sarti...	2021	TATA BUSANA	Perempuan	Wirausaha
8	Yulia sari	2021	TATA BUSANA	Perempuan	Bekerja
9	NIKEN ARISNA	2020	TATA BUSANA	Perempuan	Kuliah
10	Ricki syahputra	2021	Teknik Komp...	Laki Laki	Bekerja
11	Ike Padillah	2021	Teknik Komp...	Perempuan	Kuliah
12	Adilla Azzahra	2020	TATA BUSANA	Perempuan	Wirausaha
13	SAQINA ARD...	2020	TATA BUSANA	Perempuan	Wirausaha
14	HILDA TRIAN...	2021	TATA BUSANA	Perempuan	Wirausaha
15	Win Gomgo...	2019	Teknik Komp...	Laki Laki	Kuliah

Figure 5. Data transformation

Clustering data using RapidMiner Studio

The next step is to ask the operator to retrieve the data from the imported location and ensure that the data is ready to be processed to the next step.

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

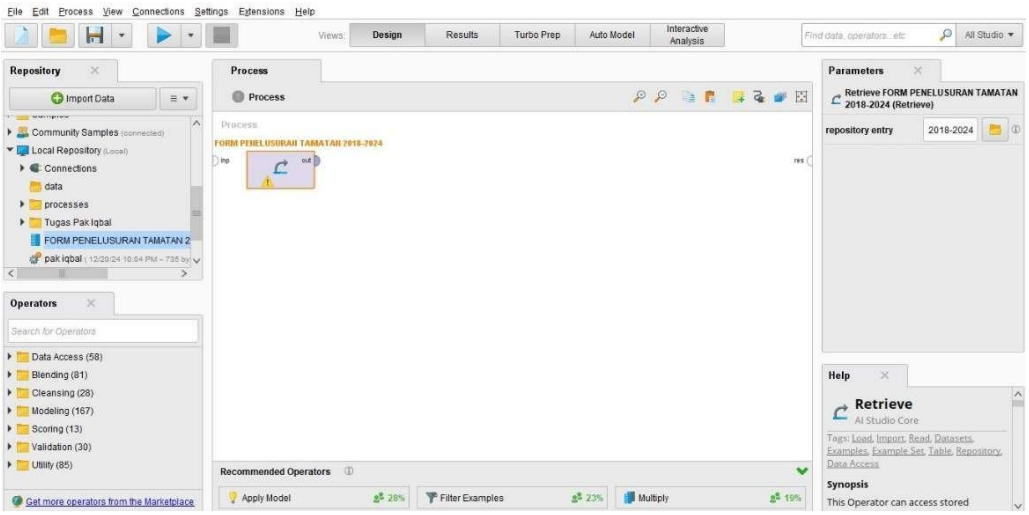


Figure 6.Data Retrieval Operators

The next stage is to build a clustering model using the k-means method, which will produce a number of clusters according to the analysis needs. This model aims to group data based on certain similarities or characteristics, so that it can support the interpretation process and subsequent decision making.

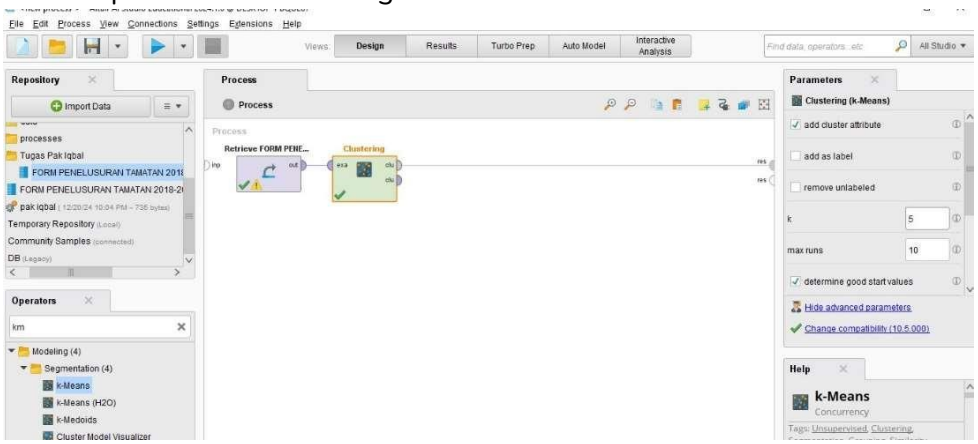


Figure 7.K-Means Operator

Next, this stage explains the results of the Cluster Model Operator, which produces a division into 4 clusters. From a total of 286 datasets analyzed, each cluster has a different number of members: Cluster 0 consists of 134 members, Cluster 1 contains 143 members, Cluster 2 has 0 members and Cluster 3 has 9 members. These results provide an overview of the distribution of data in each cluster based on the characteristics that have been analyzed previously.

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

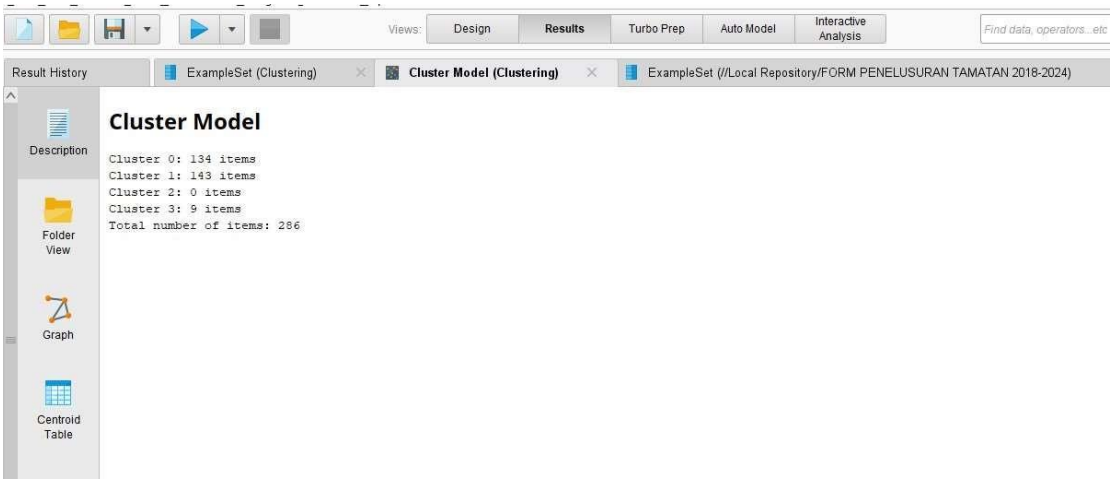


Figure 8. *Cluster Model*

CONCLUSION

This study successfully identified the career patterns of SMKN 1 Stabat graduates using the K-Means algorithm. From the results of the analysis, there are four main groups of career patterns, namely graduates who work in the industrial sector, continue their education, become entrepreneurs, and those who are not yet working. Each group has unique characteristics that can be a reference for schools in formulating curriculum and learning development strategies. The application of the K-Means algorithm has proven effective in grouping tracer study data and providing useful insights for strategic decision making. With the results of this study, SMKN 1 Stabat is expected to be more responsive to the needs of the world

REFERENCES

[1] Ahmad, A. et al. (2022). "Data Mining Techniques in Educational Research."
[2] Ojs, A. (2017). Widyowati, Dyah Penyelarasan Kurikulum Pendidikan Vokasi
[3] Wardina, U. V., Jalinus, N., & Asnur, L. (2019). Kurikulum pendidikan vokasi pada era revolusi industri 4.0. *Jurnal pendidikan*, *20*(1), 82-90.
[4] Sembiring, S. N. B., Winata, H., & Kusnasari, S. (2022). Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, *1*(1), 31-40.
[5] Fajriyah, N., Setiawan, W., Dewi, E., & Duha, T. (2022). Implementasi teknologi big data di era digital. *Jurnal Informatika*, *1*(1), 1-7.
[6] Sulistiyawati, A., & Supriyanto, E. (2021). Implementasi Algoritma K-means Clustring dalam Penetuan Siswa Kelas Unggulan. *Jurnal Tekno Kompak*, *15*(2), 25-36.
[7] Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. *Pattern recognition*, *36*(2), 451-461. Fayyad, U. M. et al. (1996). "From Data Mining to Knowledge Discovery in Databases."
[8] Setiawan, A. et al. (2020). "Implementasi Clustering pada Data Pendidikan."
[9] Smith, T. et al. (2019). "Silhouette Analysis for Clustering Validation."

Journal Of Data Science
Volume 3, Number 01, 2025, DOI 10.58471/Science.v3i01
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

[10] Witten, I. H., et al. (2016). *Data Mining: Practical Machine Learning Tools.*
[11] Purwanto, H. (2021). "Penerapan Data Mining pada Data Tracer Study."
[12] Suhartono, T. (2019). "Pendidikan Vokasi dan Tantangan Dunia Kerja."
[13] Tan, P. N., et al. (2005). *Introduction to Data Mining.*
[14] Fajriyah, S. (2022). "K-Means Clustering untuk Analisis Pola Karir."
[15] Rizal, D. (2020). "Tracer Study untuk Pendidikan Kejuruan di Indonesia."
[16] Hartigan, J. A., & Wong, M. A. (1979). "Algorithm AS 136: A K-Means Clustering Algorithm."
[17] Suyanto. (2018). *Machine Learning Dasar Teori dan Aplikasi.*
[18] Susanto, D. (2020). "Evaluasi Pendidikan Berbasis Data Mining."
[19] Nurhayati, N. (2021). "Strategi Pengembangan SMK Berbasis Tracer Study."
[20] Wahyudi, E. (2019). "Implementasi K-Means untuk Data Pendidikan Tinggi."